TOWARDS THE HOLISTIC SPATIALIZATION OF MULTIPLE SOUND SOURCES IN 3D, IMPLEMENTATION USING AMBISONICS TO BINAURAL TECHNIQUE

Coralie Diatkine *Stéphanie Bertet* *Miguel Ortiz*

Independent Researcher,
67, avenue Faidherbe,
93100 Montreuil, France
cdiatkine@coraliediatkine.eu

Queen's University Belfast,
Sonic Arts Research Centre,
Belfast, UK
s.bertet@qub.ac.uk

Goldsmiths, University of London,
Department of Computing
London, UK
m.ortiz@gold.ac.uk

ABSTRACT

This abstract describes a modular tool, dedicated to the real time spatialization of multiple sources in three dimensions, based on a mixed Ambisonics or Higher Order Ambisonics (HOA) to binaural technique, coupled with an interface that allows to position sound sources using free-hand gestures in a visual 3D environment. It is implemented in the real-time programming environment Max/MSP.

1. INTRODUCTION

Different approaches to binaural synthesis exist. The Ambisonics to binaural technique allows the rendering of a sound field representation in HOA by binaural synthesis of virtual loudspeakers, through headphones. Ambisonics is a spatialisation technique based on spherical decomposition of the sound field. The sound field is encoded to ambisonic components (spherical harmonics) and decoded on loudspeakers configuration. The more ambisonics components used (the higher the ambisonic order), the more precise the reproduction. It is based on a uniform distribution of speakers, around a listening position. However, as the spatial resolution increases, its conventional application makes it inaccessible outside specialized studios. Binaural synthesis is convenient to render a unique, static source, but inefficient when applied to multiple moving sources. The process, based on Head Related Transfer Function (HRTF) interpolation and their convolution with the sources yields imprecise results, at a high computational cost [1]. A mixed ambisonics to binaural technique enables the rendering of the ambisonics representation of a sound field, encoded and decoded in Ambisonics through virtual speakers, and then synthesized as a binaural audio stream [2]. This allows to increase the number of sources, while improving the binaural rendering in 3D, whether sources are static, or dynamically relocated. Using a gesture controller and applying it to 3D allows a straightforward and intuitive exploration of the relationship between sound, space and interpretation of movement.

2. REPRODUCTION LAYOUT DESIGN

Designing an ideal 3D loudspeaker layout is a non-trivial task [3]. The number of speakers *NS* is expressed proportionally to the order *N* by $NS = (N+1)^2$. *NS* increases exponentially with the precision of the representation, with a large number of speakers below and overhead the listener. To provide the best possible HOA representation in a generic context, we chose to use conventional uniform speaker layouts. For each order, a periphonic layout was computed with the Matlab 3LD library [4]. The number of speakers and their coordinates were made as close as possible to the ideal *NS*, and HRTFs measurements.

Available HRTF databases are not necessarily compliant with 3D HOA speaker configurations therefore, a trade-off had to be made between HRTFs resolution, even speaker arrays design, and the computational power required by a joint HOA representation and binaural synthesis with a large collection of HRTFs. Following informal listening, the Ircam's database [6] yielded excellent results in terms of externalization and 2D localisation accuracy. Furthermore, it is fully compatible with a 3rd order speaker layout, and close to other orders' layouts. Higher orders require *ad hoc* HRTFs. Concomitantly with growing computational power the use of high-resolution databases should help reach increasingly precise sound field representations with headphones.

3. SOFTWARE MODULES

A series of independent modules that divides the task of ambisonics encoding/decoding, binaural rendering and gestural control of sound source positioning are implemented. The tool's architecture is presented in figure 1.

4.1. Ambisonics representation module

The maximum number of encoded sources is set to 250, the inputs being activated / deactivated in real time by the creation or destruction of sound sources, without extra CPU consumption. The sphere radius and distance encoding can naturally be modified. The decoding parameters and speakers coordinates are called when selecting an order. The encoder uses ICST plugins [8] whereas the decoder is the SARCoder, VST plugin developed at SARC. The three well known decoding options are available : standard, max_{rE}, and in-phase.

4.2. Binaural synthesis module

The binaural synthesis of the ambisonic representation is executed by a module that ensures the convolution of the audio signal emitted by a virtual speaker, with a left or a right HRTF. To optimize the CPU usage, this module was designed as a *[poly~]*, that is duplicated into as many

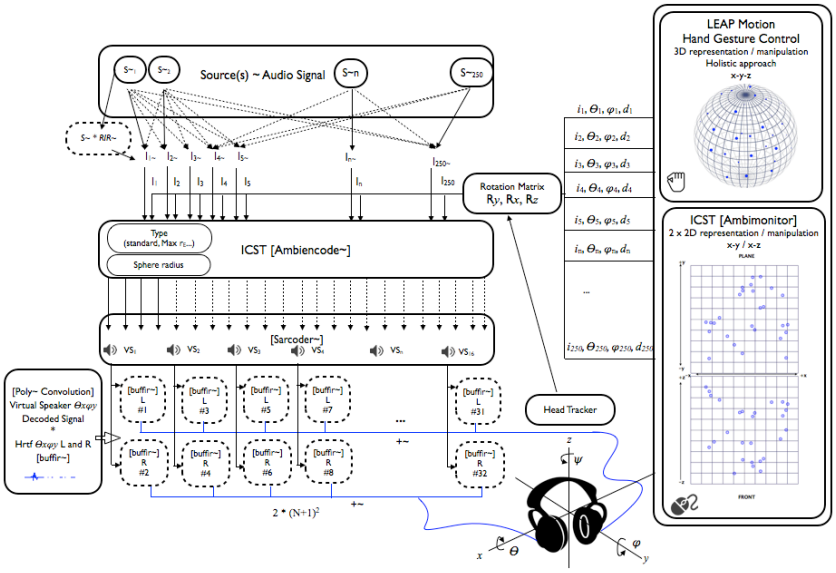

Figure 1: Overview of the different modules

instances as necessary. The HRTF file is selected according to the ambisonics order and speaker configuration. It is stored in a buffer to be convolved with the incoming audio signal. The resulting signal is then routed to the left or right audio headphone input.

4.3. Sources positioning module

The concept of source denotes the result of the interdependency between a sound and its spatial location. ICST's *Ambimonitor* is a graphic interface used for the manual or automated positioning of points in the horizontal and vertical planes. Points can be created and deleted on the fly, within a flexible spatial scope. This operation automatically establishes audio and data connections to the encoder. Each point gets a random position expressed in Cartesian or polar coordinates. The number of points is independent from the number of signals, which enables one-to-one connections, multilayered sources, or decorrelated signals [7]. Points can be deleted and retrieved to explore the effect of density variations. These features contribute to the exploration of different sources-space relationships.

Additionally, we have worked on incorporating an experimental module to allow for a more intuitive interaction to position sound sources in 3D space. The module makes use of the LEAP Motion Controller [8], an affordable interface that allows for hand gesture control in Human Computer Interaction contexts. Each sound source defined in the *Ambimonitor* is displayed in a virtual 3D environment, where the representation of the user's hands allows to hover over the sound sources. A basic « pinch » gesture allows the user to grab a sound source and freely move it in real time within the 3D space. Releasing the pinch 'drops' the sound source at the current 3D position. This module focuses on a holistic approach to interacting and exploring sound source positioning using free-hand gestures, as a complement to more deterministic modes of interaction where spatial coordinates are specified as provided by the *Ambimonitor*.

5. FURTHER DEVELOPMENT

Reverberation is a major cue in distance assessment [9] and sources localization accuracy. Therefore a robust solution needs to be implemented, using natural or simulated Room Impulse Responses. Additional work planned will be the addition of a head tracker for a higher sound field stability and further improvements to the free-hand gestural controller for the real time manipulation of sound sources position.

6. REFERENCES

[1] V. Pulkki, M. Karjalainene, J. Huopaniemi, "Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model", Journal of Audio Engineering Society, 47(4), April 1999, pp. 203-2017

[2] M. Noisternig, T. Musil, A. Sontacchi, R. Höldrich, "A 3D real time rendering engine for binaural sound reproduction", Proceedings of the 2003 International Conference on Auditory Display, Boston, 6-9 July 2003

[3] A. J. Heller, E. M. Benjamin, "The ambisonics decoder Toolboox: Extensions for Partial-Coverage Loudspeaker Arrays", Linux Audio Conference 2014, ZKM, Karlsruhe, Germany. Keynote : http://goo.gl/XxsRkL

[4] http://flo.mur.at/writings/3LD-iem-report.pdf/view

[5] http ://recherche.ircam.fr/equipes/salles/listen/index.html

[6] J. Schacher, "Seven years of ICST Ambisonics Tools for Max/MSP – A brief report", Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics , Paris, May 6-7, 2010

[7] Kendall, G., The decorrelation of audio signals and its impact on spatial imagery, Computer Music Journal Vol. 19, No. 4 Winter, 1995, pp. 71-87

[8] https://www.leapmotion.com

[9] S. H. Nielsen, "Auditory distance perception in different rooms", J. Audio Eng. Soc. Vol. 41, No 10, October 1993